LEC/LADE/LEC-DER/LADE-DER/LEC-PERIODISMO
CURSO 2007/08
EXAMEN DE ESTADÍSTICA I
16 de Febrero de 2008
TIEMPO MÁXIMO = 3h

1. **(3 puntos).** La empresa "Inmobiliaria S.A" gestiona el arrendamiento de inmuebles en la Costa Blanca. En particular, para los pisos dedicados al alquiler de temporada, se desea estimar el tiempo medio de duración de los alquileres (en días) y su variabilidad. Realizado un muestreo de 16 alquileres de dichos pisos, se obtuvieron los siguientes resultados:

   52; 36; 65; 22; 35; 60; 27; 38; 15; 90; 50; 45; 31; 60; 15; 83

   Además se puede ver que $\sum_{i=1}^{16} x_i = 724$ y $\sum_{i=1}^{16} x_i^2 = 40232$. Se pide,

   a) **(1 punto).** Precisando los supuestos necesarios que han de asumirse, construir un intervalo de confianza del 90% para la media.
   Solución: Para construir el intervalo de confianza debemos suponer normalidad y m.a.s. Así, el intevalo para la media μ, con desviación típica desconocida σ viene dado por $\left[\bar{x} \mp t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right]$. *En este caso,* $\bar{x} = 45.25, s = 22.32, t_{n-1,\alpha/2} = t_{15,0.05} = 1.753$, *y el intervalo viene dado por* $[35.47, 55.03]$.

   b) **(1 punto).** Precisando los supuestos necesarios que han de asumirse, construir un intervalo de confianza del 95% para la desviación típica. De acuerdo a estos resultados, ¿se admitirá (con un nivel de significación del 5%) una desviación típica poblacional de 25 días?
   Solución: Volvemos a suponer normalidad y m.a.s. En este caso, el intervalo de confianza al nivel $1-\alpha$ *para* σ^2 *es* $\left[\frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}}\right]$. *En este caso,* $\left[\frac{15\cdot 498.07}{\chi^2_{15;0.025}}, \frac{15\cdot 498.07}{\chi^2_{15;0.975}}\right] = [271.68, 1193.46]$. *Y por tanto, el intervalo para la desviación* σ *es* $[16.48, 34.55]$. *Los valores de la* χ^2 *son* $27.5, 6.26$. *Como el valor 25 pertenece al intervalo de confianza del* 95%, *no tenemos evidencia estadística para rechazar una desviación típica poblacional igual a 25 con nivel de significación del* 5%.

   c) **(1 punto).** Con los mismos datos y asumiendo un 1% de nivel de significación, ¿hay evidencia muestral para rechazar la hipótesis de que la desviación típica poblacional es de 25 días?

   Solución 1: *mediante justificación lógica*
   ya que (i) el contraste

$$H_0 : \sigma = \overbrace{25}^{\sigma_0} \qquad vs \qquad H_A : \sigma \neq 25$$

a nivel 1% es equivalente a comprobar si el valor $\sigma_0 = 25$ pertenece al intervalo de confianza al 99% para σ; (ii) el interval al 95% para σ está contenido en el intervalo al 99% para σ; (iii) y debido al apartado b); no hay suficiente evidencia empírica al 1% para rechazar la nula y se puede concluir que σ es 25.

Solución 2: *Realizamos el test*
Hipótesis nula y alternativa:

$$H_0 : \sigma = \overbrace{25}^{\sigma_0} \qquad vs \qquad H_A : \sigma \neq 25$$

Estadístico:

$$\chi^2 = \frac{(n-1)}{\sigma_0^2} s^2 = \frac{15}{25^2} 498.0667 = 0.024(498.0667) = 11.9536 \approx 11.95$$

Decisión y conclusión usando el p-valor: $11.04 < 11.95 < 14.34$

$$P(\chi^2 \geq 11.95) \in (0.5, 0.75)$$

por tanto al 1% no rechazamos la nula y concluimos que $\sigma = 25$
Solución 3: *Repetimos el procedimiento del apartado b) y obtenmos un nuevo intervalo de confianza pero esta vez al* 99%*. Ahora los valores de la* χ^2 *son* $32.8, 4.6$*, el intervalo para* σ^2 *es [227.7759,1624.1] y para* σ *[15.09,40.30]. Así pues la conclusión es la misma que la del apartado b).*

2. **(2 puntos).** Cierta publicación técnica dio a conocer los resultados de un análisis del peso (mg) de calcio en cemento estándar y en cemento contaminado con plomo. Los niveles bajos de calcio indican que el mecanismo de hidratación del cemento queda bloqueado y esto permite que el agua ataque varias partes de una estructura de cemento. Al tomar 10 muestras de cemento estándar, se encontró que, en ciertas unidades, el peso promedio de calcio es de 90 con una desviación estándar de 5; los resultados obtenidos con otra muestra independiente de tamaño 15 de cemento contaminado con plomo fueron de 87 en promedio con una desviación estándar de 4. Supóngase que el peso del calcio está distribuido de manera normal y que las dos poblaciones normales tienen la misma desviación estándar.

   a) **(1.5 puntos).** Hallar un intervalo de confianza del 95% para la diferencia entre los pesos promedio de calcio de los dos tipos de cemento.
   Solución: Sean X = *peso del cemento estándar e* Y = *peso del cemento contaminado. Entonces según el enunciado,* $X \sim N(\mu_1, \sigma), Y \sim N(\mu_2, \sigma)$*. Buscamos un intervalo para* $\mu_1 - \mu_2$*, es decir, para la diferencia de dos medias de poblaciones normales e independientes. Este intervalo (en este caso, las varianzas son desconocidas pero iguales) viene dado por* $\left[\bar{x} - \bar{y} \mp t_{n_1+n_2-2;\alpha/2} s_p \sqrt{1/n_1 + 1/n_2}\right]$.
   Las cuasivarianzas las obtenemos a partir de las desviaciones muestrales, $s^2 =$

$n\sigma^2/(n-1)$.En este caso, puede comprobarse que $s_x^2 = 27.78, s_y^2 = 17.14$,$s_p = \frac{9\cdot 27.78+14\cdot 17.14}{23} = 21.3$ y por tanto, $s_p = 4.61$. Además $t_{23,0.025} = 2.069$, y el intervalo final es $[-0.8939, 6.8939]$.

b) **(0.5 punto).** ¿Se puede rechazar la hipótesis de igualdad de medias teniendo en cuenta el intervalo de confianza del apartado anterior?
Solución: Como el 0 está incluido en el intervalo anterior, concluimos que no hay evidencia muestral para rechazar la hipótesis de igualdad de medias (al 5%).

3. **(3 puntos).** Dada una muestra aleatoria simple de n observaciones de una variable aleatoria continua con función de densidad

$$f(x) = \begin{cases} \theta(1-x)^{\theta-1} & \text{si } 0 \leq x \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

siendo θ un cierto parámetro estrictamente mayor que 0, se pide:

a) **(1.5 puntos).** Obtener el estimador máximo verosímil de θ.
Solución: Si la muestra aleatoria simple es $(x_1, x_2 \ldots x_n)$, y teniendo en cuenta la independencia de las x_i, se tiene que la función de verosimilitud viene dada por

$$f_\theta(x_1, x_2 \ldots x_n) = \prod_{i=1}^{n} \theta(1-x_i)^{\theta-1} = \theta^n \prod_{i=1}^{n} (1-x_i)^{\theta-1}.$$

Así, la función soporte toma la forma

$$L(\theta, x_1, x_2 \ldots x_n) = n ln\ \theta + (\theta - 1)\sum_{i=1}^{n} ln\ (1-x_i).$$

Derivando con respecto a θ e igualando a cero, obtenemos

$$0 = L'(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} ln\ (1-x_i),$$

con lo que la raíz es $\theta = -\frac{n}{\sum_{i=1}^{n} ln\ (1-x_i)}$.
Para comprobar si esta raíz es en efecto el estimador buscado, hemos de ver que el valor de la segunda derivada en ella es negativa (condición de máximo):

$$\frac{\partial^2 L}{\partial \theta^2}(\theta) = -\frac{n}{\theta^2}.$$

Como este último número siempre es negativo, concluimos que $\hat{\theta}_{MV} = -\frac{n}{\sum_{i=1}^{n} ln\ (1-x_i)}$.

b) **(1 punto).** Calcular la varianza asintótica (estimada) de dicho estimador.
Como la varianza asintótica es el inverso del valor en $\hat{\theta}_{MV}$ *de* $-\frac{\partial^2 L}{\partial \theta^2}(\theta)$, *se tiene que* $Var(\hat{\theta}_{MV}) = \frac{\hat{\theta}^2_{MV}}{n} = \frac{n}{[\sum_{i=1}^{n} ln(1-x_i)]^2}$.

c) **(0.5 puntos).** Obtener los valores concretos del estimador máximo verosímil y la varianza asintótica (estimada) para una muestra aleatoria simple cuyos valores son 0.2, 0.5 y 0.8.
Tomando valores en los elementos muestrales dados, se obtiene que en este caso concreto $\hat{\theta}_{MV} = 1,188$ *y* $Var(\hat{\theta}_{MV}) = 0,4703$.

4. **(2 puntos).** Después de las Navidades, la Oficina de Atención al Consumidor (OAC) ha sufrido un aumento considerable de llamadas exponiendo quejas. Para un estudio de calidad se quiere valorar si la variable *tiempo de espera hasta que la queja es atendida* (en días) puede ajustarse por una distribución Normal.

Para ello, la OAC toma una muestra aleatoria simple de $n = 200$ “tiempos de espera”, y obtiene:

Tiempo de espera	0	1	2	3	4	5	6	7	8	9	10
Frecuencia observada	4	6	7	31	45	50	34	10	4	6	3

La siguiente tabla muestra los valores esperados obtenidos si la variable ”tiempo de espera” siguiera una distribución Normal.

Tiempo de espera	menos de 3	3	4	5	6	7	8 o más
Frecuencia esperada	20	30	40	40	30	20	20

a) **(1 punto).** Calcular el valor del estadístico y los grados de libertad al realizar el test de bondad de ajuste de la χ^2.
Solución: Tenemos 7 clases, luego
$\mathcal{X}_c^2 = \sum_{i=1}^{7} \frac{(E_i - O_i)^2}{E_i} = 11.53$
Los grados de libertad son 4, $\mathcal{X}^2_{7-2-1} = \mathcal{X}^2_4$, *ya que tenemos 7 clases y desconocemos los 2 parámetros poblacionales de la Normal.*

b) **(1 punto).** Determinar si existe evidencia estadística (con un nivel del 5%) para rechazar la hipótesis de normalidad.
El contraste planteado es
$H_0 : X \sim \mathcal{N}(\mu, \sigma^2)$
$H_1 : X \nsim \mathcal{N}(\mu, \sigma^2)$

El valor crítico a partir del cual se rechaza la hipótesis es 9,49 (Chi-cuadrado con 4 grados y alpha 0,05) Como $11,59 > 9,49$ *rechazamos la hipótesis de normalidad.*

NOTA: Las notas de cada grupo serán publicadas por cada profesor el jueves 21 de Febrero en Aula Global con las correspondientes instrucciones para solicitar revisión .